

The effects of dosage and duration of auditory training for older adults with hearing impairment

Larry E. Humes,^{a)} Dana L. Kinney, Sara E. Brown, Anna L. Kiener,
and Tera M. Quigley

*Department of Speech and Hearing Sciences, Indiana University, 200 South Jordan Avenue,
Bloomington, Indiana 47405*

*humes@indiana.edu, danakin@indiana.edu, seb@indiana.edu, akiener@indiana.edu,
tmquigle@indiana.edu*

Abstract: This study explored a range of training dosages and durations for a word-based auditory-training procedure for older adults with hearing impairment. Three groups received a different “dose”: 2x/week; 3x/week; no training. Fifteen training sessions comprised a “cycle” which was repeated three times for each dosage. Groups that completed training performed significantly better than controls for speech-in-noise materials included in the training regimen, with no significant difference observed between the 2x or 3x/week training groups. Based on these results, as well as prior literature on learning theory, training 2x or 3x/week for 5–15 weeks appears to be sufficient to yield training benefits with this training regimen.

© 2014 Acoustical Society of America

PACS numbers: 43.71.Lz, 43.71.Ky, 43.66.Sr [SGS]

Date Received: April 29, 2014 **Date Accepted:** July 9, 2014

1. Introduction

Difficulty understanding speech, especially in background noise, is one of the most common complaints of older hearing-impaired adults (OHI). Given the presence of hearing loss in speech-communication circumstances of less than optimal audibility, a better than normal speech-to-noise ratio (SNR) is often required to achieve performance that is similar to that of young normal-hearing adults.¹ Although hearing aids do seem to provide significant and long-term benefits, aided performance often remains less than ideal, especially in the presence of competing speech. Recent work in this laboratory has explored ways to improve the ability of older adults to understand amplified speech for a given SNR. Our focus in previous auditory training studies was placed on a novel approach to intervention which included the following features: (1) it was word based, which placed the focus on meaningful speech rather than on sublexical sounds of speech, such as individual phonemes; (2) it was based on closed-set identification of words under computer control, which enabled automation of presentation, scoring, and feedback; (3) multiple talkers were included in training, which facilitated generalization to novel talkers; (4) it was conducted in noise, the most problematic listening situation for older adults; and (5) both auditory and orthographic feedback associated with the correct and incorrect responses was provided to the listener after incorrect responses. These orthographic representations of the auditory stimulus reinforce the linkage between the acoustic representation of the stimulus in the impaired periphery and the intact representation of the meaningful speech stimulus in the mental lexicon.

Through a series of laboratory studies evaluating this word-based auditory training protocol,^{2–5} we have learned the following: (1) OHI listeners could improve their open-set recognition of words in noise after following the training regimen; (2)

^{a)} Author to whom correspondence should be addressed.

training generalized to other talkers saying the same trained words, but only slight improvements (7% to 10%) occurred for new words, whether spoken by the talkers heard during training or other talkers; (3) improvements from training, although diminished somewhat by time, were retained over periods as long as 6 months (maximum retention interval examined to date); (4) similar gains were observed when the feedback was either entirely orthographic (displaying correct and incorrect responses on the computer screen) or a mix of orthographic and auditory (re-hearing the correct and incorrect words after incorrect responses), but not when the feedback was eliminated entirely;² and (5) the effects of the word-based training were transferred to novel sentences and talkers when the sentences were composed primarily of words used during training.⁵ The methodological details varied from study to study, but common features included the use of closed-set identification of vocabularies of 50 to 600 words spoken by multiple talkers in background noise during training. The specific set of words and talkers varied across studies. The assessment of training effectiveness for a given study was then conducted with the stimuli used in training, as well as additional words and sentences spoken by both the same talkers used in training and by novel talkers.

The focus of the present study was to examine a range of possible training dosages and durations for this word-based auditory training system in OHI listeners. Two doses of training were investigated: Participants in the treatment groups completed training 2 days a week for 7.5-week cycles (2x/week) or 3 days a week for 5-week cycles (3x/week). In a given cycle, the total number of sessions of training and the number of stimulus exposures were the same across dosages. These treatment groups completed three cycles of training, with post-training evaluations conducted after each cycle to examine the effect of training duration.

2. Participants

Fifty-five participants (26 F; 29 M) were randomly assigned to one of three groups: A 2x/week training group ($N = 16$), a 3x/week training group ($N = 19$), and a no-training control group ($N = 20$). Participants were between 61–79 yr old ($M = 71$ y), were native English speakers, passed the Mini-Mental State examination⁶ (score > 24), and had sloping, mild to moderately severe sensorineural hearing loss. Hearing aid use was not used as an inclusion or exclusion criterion. However, 11 participants (four in the 2x/week group, three in the 3x/week group, and four in the control group) reported monaural or binaural hearing aid use which ranged in duration from 6 months to 14 years. Individuals who met selection criteria were blocked into a low, medium, or high performing (L, M, H) subgroup based on their baseline aided speech understanding performance on the Central Institute for the Deaf (CID) Everyday Sentences test.⁷ Baseline performance of $< 50\%$ was the criterion for the L subgroup, 51% – 65% for the M subgroup, and $> 65\%$ for the H subgroup. A pre-generated computer randomization of group assignment was created for each performance subgroup using a sequence of random triplet selections without replacement. This process ensured random assignment, but also ensured roughly equal numbers of low-, medium- and high-performing subjects in each group. It also allowed for sequential enrollment into the study, rather than requiring pre-identification of the entire pool of subjects prior to randomization. It is well known that improvements during training regimens are impacted by initial baseline performance such that those with lower initial scores tend to show larger improvements than those with higher initial scores. There are a variety of factors that may influence initial scores, including age, peripheral-auditory, central-auditory, and cognitive factors, but the key variable to balance across groups is the distribution of low, medium, and high baseline scores. Independent-sample t tests showed that there were no significant differences ($p > 0.05$) between the three treatment groups in terms of hearing loss or prior hearing aid use. There was a small difference ($p = 0.03$) in age between the groups with the 2x/week group being slightly younger ($M = 67$ years) than the other two groups ($M = 71$ years).

3. Stimuli and materials

Five types of speech materials were used in the evaluation and training portions of this study: (1) frequent words; (2) frequent phrases; (3) re-recordings of selected sentences from the Veterans Administration Sentence Test (VAST);⁸ (4) the Auditec recording of CID Everyday Sentences;⁷ and (5) closed-set Hagerman-format sentences⁹ composed of frequently occurring words. Materials were recorded by various combinations of six adult speakers (Talkers A, B, C, D, E, and F) of the Midland North American English dialect. Frequent word materials were recorded by three male (A, B, and C) and three female (D, E, and F) talkers. The frequent phrases and modified VAST sentences were recorded by two of the male (B and C) and two of the female (D and E) talkers, and the Hagerman-format sentences were recorded by two of the male (A and B) and two of the female (D and F) talkers. All stimuli, with the exception of one set of Hagerman-format sentences, were presented in the two-talker version of International Collegium for Rehabilitative Audiology (ICRA) envelope-modulated or vocoded noise.¹⁰ The remaining set of Hagerman-format sentences was presented in the two-talker British English babble that was used to derive the ICRA noise.¹¹ The use of meaningful speech increased the informational masking for this condition compared to the envelope-modulated vocoded ICRA noise. Details regarding the preparation and recording of the first four speech materials are available in Humes *et al.* (2009).⁵

One modification to the training protocol used in past studies from this laboratory is the addition of closed-set sentences; specifically, the Hagerman-like sentences. The approach taken by Hagerman⁹ was to constrain all the sentences to the following sequence: (Proper name) (verb) (number) (adjective) (noun-plural). This sentence format, now referred to as the Matrix Test, is available for speech perception testing in a variety of languages.¹² Sets of 250 naturally produced “Hagerman” sentences have been recorded by our laboratory by four talkers (two male and two female). Pilot testing using open-set recognition was completed for four subjects to establish performance-versus-signal-to-noise-ratio functions for these new materials. As in Humes *et al.* (2009),⁵ recognition performance was measured with the two-talker ICRA noise as competition.

As noted, we incorporated the use of two competing talkers in the initial pilot experiments as an additional source of competition and as an additional bridge to the real-life situation of listening to one talker in the presence of competing talkers. The two-talker competition was only used for one set of the Hagerman-like sentences, which included sentences recorded by all four talkers. We obtained the digital recordings of the aforementioned two talkers used to generate the noise-vocoded competition (ICRA noise): A male and a female talker, speaking (in British English) on various everyday topics.¹¹

With the incorporation of the stimulus changes noted above, the current study focused on the appropriate dose and duration of the training regimen. A design criterion imposed on the training is that the hearing aid wearer must be able to realize substantial benefits from amplification during the initial trial period. Although two-thirds of the states in the United States require a trial period of at least 30 days, many hearing-aid manufacturers willingly extend this to 60–90 days. Thus, the training program was dosed within a range of total durations of 1 to 4 months to have the most significant impact on clinical practices. Two dose regimens were examined as follows: (1) training 2 days a week for 7.5-week periods; (2) training 3 days a week for 5-week periods. Each training cycle was repeated three times to examine duration effects. Post-training evaluations occurred on a separate day following each training cycle. A control group completed only the baseline and subsequent “post-training” testing at the same temporal intervals as those in the 3x/week training group. The control group was included to assess the learning or familiarization effects that may occur due to the requirement for repeated assessments in the experimental groups.

4. Equipment and calibration

Stimuli were spectrally shaped based on each participant’s pure tone thresholds to simulate a well-fit hearing aid that optimally restored audibility of the speech

spectrum through at least 4000 Hz and resulted in the speech signal being at least 10 dB above threshold at 5000 Hz. Additional details regarding this approach to spectral shaping can be found in the study by Burk and Humes (2008).⁴ Typically, after spectral shaping, the overall sound pressure level (SPL) of the steady-state calibration stimulus would be approximately 100 to 110 dB SPL. The following SNRs were used during assessment and training: 2 dB for frequent words; −8 dB for frequent phrases, VAST sentences, and CID Everyday sentences; −9 dB for Hagerman-like sentences in ICRA-derived noise; and −6 dB for Hagerman-like sentences in two-talker British English babble. At first glance, these SNRs seem to be very severe. However, it should be kept in mind that these SNRs are defined on the basis of the steady state calibration noise whereas the actual competition was the fluctuating two-talker vocoded ICRA noise, or the two-talker British-English babble in the case of the second set of Hagerman-format sentences. These SNRs were selected to place initial baseline performance for most listeners in the range of 40%–70% correct; well below ceiling, but not too low to be overly frustrating to the listeners. All testing and training were completed in sound-treated test booths meeting ANSI (1999)¹³ ambient noise standards for threshold testing under earphones. Eight listening stations were housed within the sound-treated booths, each equipped with a 17-inch liquid crystal display flat-panel touch screen monitor. All testing was monaural. All responses during closed-set identification training were made via the touch screen. For word-based training, 50 stimulus items were presented on the screen at a time, in alphabetical order from top to bottom and left to right, with a large font that was easily read (and searched) by all participants. Additional features of the word-based training protocol have been described in detail in Humes *et al.* (2009).⁵ With regard to the phrase-based training, closed sets of 11 or 12 phrases were displayed on the computer screen at a time and the listener's task was to select the phrase heard. The training for phrases was otherwise identical to that for words. In contrast to previous studies,^{3–5} feedback for each trial was solely orthographic with no opportunity to listen again to a missed word.

5. Procedures

The pre-training baseline testing consisted of the presentation of (1) 150 frequently occurring words spoken by six talkers; (2) 94 frequent phrases spoken by four talkers for a total of 434 words scored; (3) 20 CID Everyday Sentences (Lists A and B) digitized from the Auditec CD with 100 keywords scored; (4) 50 VAST sentences produced by four talkers for a total of 200 sentences with 600 keywords scored; and (5) 100 Hagerman-format sentences produced by four talkers. The order of presentation for the post-training assessments was identical to pre-training except that additional sets of 20 CID Everyday Sentences (Lists C, D, E, and F) were also presented. The rationale for including an alternate set of sentences was that there was a slight possibility that the listeners might retain some of the previous sets of CID Everyday Sentences in memory from the exposures several weeks earlier. The stimulus sequence within a given test was randomized. After pre-training evaluation, the two training groups completed three cycles of training followed by post-training evaluations after each cycle. For the 2x/week group, the duration of each cycle was 7.5 weeks with post-training evaluation completed at 7.5, 15, and 22.5 weeks re: baseline testing. The group that trained 3x/week completed 5-week cycles with post-training evaluation completed at 5, 10, and 15 weeks re: baseline testing. The control group completed only “post-training” evaluations at 5, 10, and 15 weeks re: baseline testing. The group that trained 2x/week completed training either on a Monday/Thursday or a Tuesday/Friday schedule whereas the 3x/week group trained on a Monday/Wednesday/Friday schedule. Each of these schedules made use of distributed training sessions and uniform intervals that were found to be optimal for acquisition and retention of new information in a review on learning sequencing by Cepeda *et al.* (2006).¹⁴

For each cycle of training, both training groups (2x/week and 3x/week) completed 15 training sessions that were administered on separate days. Sessions 1–6 and

10–15 were dedicated to training using the isolated frequent words and Hagerman-like sentences in ICRA-derived speech-like noise. In each of these twelve 75 to 90 min training sessions, six blocks of 50 lexical items were presented for a total of 300 lexical items, with 50 presented by each of the six talkers. The training sessions for each of these days concluded with the presentation of two blocks of 25 Hagerman-like sentences, each containing five words for a total of 250 lexical items. Sessions seven through nine consisted of 2 days of training (sessions seven and nine) with frequent phrases, half (94) of the 188 phrases presented in each session and with approximately 25% (23–24) of the phrases spoken by each of four talkers, and one day of training (session eight) with Hagerman-like sentences in two-talker British English babble. Session eight involved the presentation of eight blocks of 25 Hagerman-like sentences in two-talker British English babble using a total of 200 five-word sentences or 1000 lexical items. Post-training evaluations were completed for all groups of participants 2–3 days following each cycle completion and consisted of open-set speech recognition in noise described previously. In all training conditions in this study, during training, the listeners were only given one presentation of the training stimuli followed by correct/incorrect feedback. Unlike prior studies with this training regimen, the subjects were not allowed to hear the training stimuli multiple times in immediate succession.

For the newly incorporated Hagerman-like sentences, five columns each containing ten words appeared on the screen. A five-word sentence was presented and the listener then selected one word in each column which would highlight the selected word in blue. Once the listener selected one word from each column, he or she pressed “okay” to submit the sentence. Each correctly chosen word in the sentence became highlighted in green and filled a corresponding blank at the top of each column. If the listener made an incorrect selection, that word would be highlighted in red with the corresponding correct word highlighted in green within the same column. If any of the words selected were incorrect, the sentence was scored based on the number of correct words identified (e.g., three of five words correct), the cumulative percent-correct score visible to the subject was updated on the screen, and the next sentence was presented.

6. Results and discussion

Figure 1 shows the means and standard deviations of the three treatment groups on each of the four sets of speech materials used in training. Results are shown for baseline testing and at each of the three post-training sessions. Figure 2 shows plots of the same type, but for the two sets of speech materials not included in training (VAST, CID).

A separate mixed General Linear Model (GLM) analysis was performed on the data for each of the six speech measures with a between-subjects factor of group and a repeated-measures factor of post-training session. No significant differences ($p > 0.05$) were observed between the three groups (2x/week; 3x/week; control) for any of the six open-set speech recognition tests at baseline (pre-training). In addition, there was no significant difference in post-training evaluation performance for the two training groups (2x/week and 3x/week) and no significant interactions with group. As a result, the two training groups’ data were combined in subsequent analyses and hereafter are referred to as the “trained group.” Statistical analyses revealed a significant main effect of treatment group, showing that the trained group always performed significantly better than the control group for materials used during training (Fig. 1). There was also a significant effect observed for the main factor of post-training session and the interaction between the post-training session and group factors. *Post hoc* testing showed that this interaction was due to a significant improvement from post-training session one to post-training session two for all four sets of trained materials, but only from post-training session two to post-training session three for one set of stimuli (Hagerman-format sentences in ICRA noise).

The GLM analyses for the two sets of untrained speech materials (Fig. 2) showed no significant main effects of group and no significant interactions with group. Thus, the training in this study did not generalize to untrained sentences. For the CID

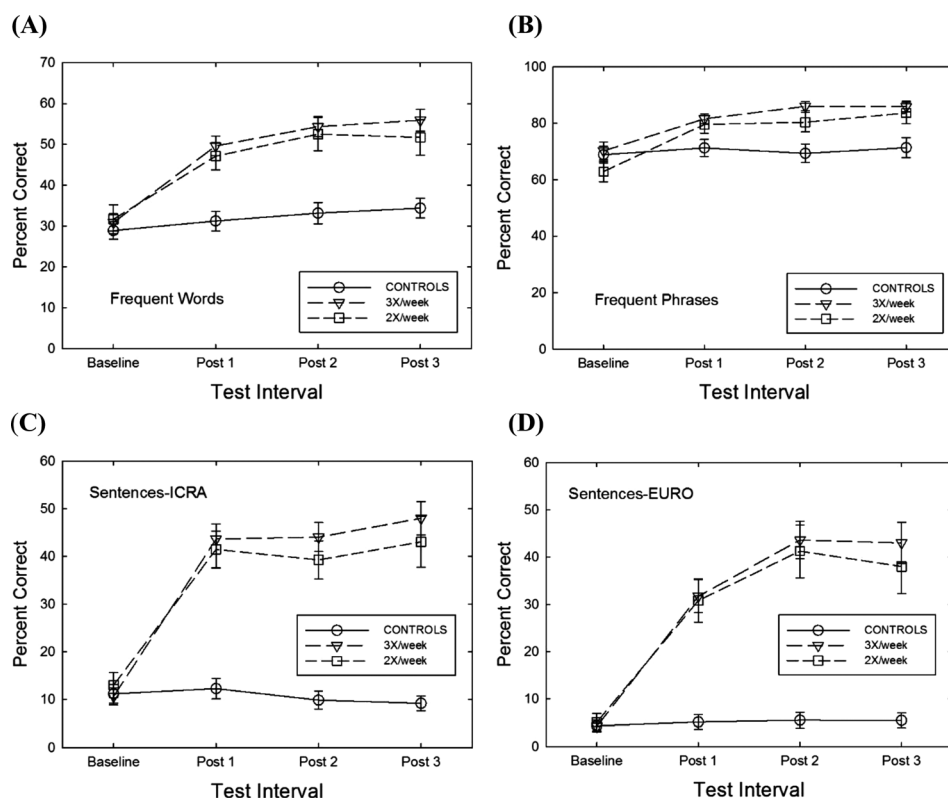


FIG. 1. Results for all three groups on the four sets of trained speech materials: (A) frequent words; (B) frequent phrases; (C) frequent-word sentences in ICRA noise; and (D) frequent-word sentences in “EURO noise” (British-accented English, two competing talkers). Error bars depict ± 1 standard error.

Everyday sentences, however, there was a significant main effect of post-training session with post-training session two yielding significantly lower performance than that at all other measurement intervals. This is clearly identifying a lack of inter-list equivalency for the lists of CID Everyday Sentences used in this study and has nothing to do with training.

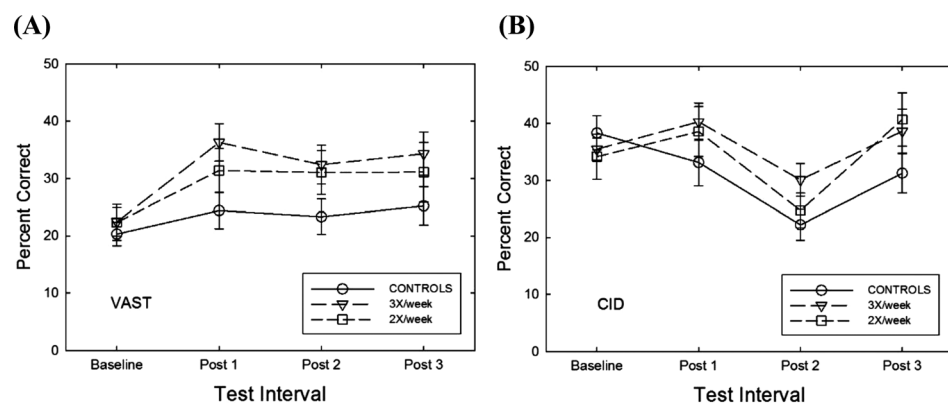


FIG. 2. Results for all three groups on the two sets of *untrained* speech materials: (A) VAST sentences; and (B) CID sentences. Error bars depict ± 1 standard error.

7. Conclusions

In general, this study, as did our earlier studies in this same area, supports the efficacy of a word-based approach to auditory training in OHI listeners. For the range of dosages and durations considered here, the results indicate that training two to three times per week and for 5 (3x/week) to 15 (2x/week) weeks would be sufficient to show significant improvements on the open-set recognition of trained speech materials in noise. The best training duration will vary with one's criterion for training completion. If one considers the magnitude of the improvement and the number of trained tests for which significant improvements were observed, then one would conclude that training was complete at post-training session two (5–7.5 weeks, depending on training frequency). If one considers training incomplete as long as any trained material showed significant improvement, then one would conclude that training was likely complete at session three (10–15 weeks) when performance increased significantly on only one of the four sets of training materials. Thus, depending on one's criterion, 5–15 weeks may be sufficient to maximize training effects. The lack of generalization to untrained talkers and sentences observed here may be due to the fact that listeners were not given the opportunity to listen again to the correct and incorrect materials in quick succession as they were in our earlier work. Future studies from this laboratory will focus on the evaluation of a home-based version of this auditory-training protocol, three times per week for 5 to 10 weeks, with the participant's own hearing aids.

Acknowledgments

This work was supported, in part, by NIDCD R01 DC010135.

References and Links

- ¹L. E. Humes and J. R. Dubno, "Factors affecting speech understanding in older adults," in *The Aging Auditory System*, edited by S. Gordon-Salant, R. D. Frisina, A. N. Popper, and R. R. Fay (Springer, New York, 2010), pp. 211–258.
- ²M. H. Burk, L. E. Humes, N. E. Amos, and L. E. Strauser, "Effect of training on word-recognition performance in noise for young normal-hearing and older hearing-impaired listeners," *Ear. Hear.* **27**, 263–278 (2006).
- ³M. H. Burk and L. E. Humes, "Effects of training on speech recognition performance in noise using lexically hard words," *J. Speech. Lang. Hear. Res.* **50**, 25–40 (2007).
- ⁴M. H. Burk and L. E. Humes, "Effects of long-term training on aided speech-recognition performance in noise in older adults," *J. Speech. Lang. Hear. Res.* **51**, 759–771 (2008).
- ⁵L. E. Humes, M. H. Burk, L. E. Strauser, and D. L. Kinney, "Development and efficacy of a frequent-word auditory training protocol for older adults with impaired hearing," *Ear. Hear.* **30**, 613–627 (2009).
- ⁶M. F. Folstein, S. E. Folstein, and P. R. McHugh, "'Mini-mental state.' A practical method for grading the cognitive state of patients for the clinician," *J. Psychiatr. Res.* **12**(3), 189–198 (1975).
- ⁷H. Davis and S. R. Silverman, *Hearing and Deafness*, 2nd ed. (Holt, Rinehart, and Winston, New York, 1960), pp. 548–552.
- ⁸T. S. Bell and R. W. Wilson, "Sentence recognition materials based on frequency of word use and lexical confusability," *J. Am. Acad. Audiol.* **12**, 514–522 (2001).
- ⁹B. Hagerman, "Sentences for testing speech intelligibility in noise," *Scand. Audiol.* **11**, 79–87 (1982).
- ¹⁰W. A. Dreschler, H. Verschuure, C. Ludvigsen, and S. Westermann, "ICRA noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment," *Audiology* **40**, 148–157 (2001).
- ¹¹D. Chan, A. Fourcin, D. Gibbon, B. Grandstrom, M. Huckvale, G. Kokkinakis, K. Kvale, L. Lamel, B. Lindberg, A. Moreno, J. Mouropoulos, F. Senia, I. Trancoso, C. Veld, and J. Zeiliger, "EUROM – a spoken language resource for the EU," in *Proceedings of 4th European Conference on Speech Communication and Technology*, Madrid, Spain (September 18–21, 1995).
- ¹²M. A. Zokoll, S. Hochmuth, A. Warzybok, K. C. Wagener, M. Buschermöhle, and B. Kollmeier, "Speech-in-noise tests for multilingual hearing screening and diagnostics," *Am. J. Audiol.* **22**, 175–178 (2013).
- ¹³ASA (2013). S3.1-1999. *American National Standard Maximum Permissible Noise Levels for Audiometric Test Rooms* (Acoustical Society of America, Melville, NY, 2013).
- ¹⁴N. J. Cepeda, H. Pashler, E. Vul, J. T. Wixted, and D. Rohrer, "Distributed practice in verbal recall tasks: A review and quantitative synthesis," *Psychol. Bull.* **132**, 354–380 (2006).